This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

10/001557

Description

SELF-ALIGNED NON-VOLATILE MEMORY CELL

CROSS-REFERENCE TO RELATED APPLICATION

This is a divisional of pending patent application Serial No. 09/727,571 filed November 30, 2000.

TECHNICAL FIELD

The invention relates to self-aligned non-volatile memory cells, and more particularly to a self-aligned non-volatile memory cell that has a high capacitive coupling ratio and has a thin and small tunneling oxide region.

BACKGROUND ART

Fig. 1 shows a sectional view of an EEPROM 100 (Electrically Erasable Programmable Read Only Memory) as depicted in Fig. 18 of U.S. Patent No. 4,833,096 which is assigned to the same assignee as the present invention. With reference to Fig. 1 of the present application, a deep n-well 23 is formed inside a p-type substrate 40, and N-channel and memory cell devices are defined. N-channel stages and field oxide are formed around the device areas. Channel stops and field oxide are formed by thermally growing a thin oxide layer, depositing a 1000-2500Å thick nitride layer and removing the nitride from non-device areas, implanting boron ions around the N-well and N-channel device areas, then driving in the boron and thermally growing oxide in the non-device areas not covered by nitride.

The process continues by implanting a first species of N-type impurity in a portion of the memory cell device area, thermally growing a first oxide layer 59, defining a window therein over the impurity implant, implanting a second species of N-type impurity into the window hole, and regrowing a thick oxide layer in the window. Next, a 2500-3400Å thick polycrystalline silicon

("polysilicon") layer is deposited, and removed with the first oxide layer to form the floating gate 71. A second oxide layer is thermally grown at a temperature of 1000°-1050° C. which ensures that this second oxide layer has a substantially uniform thickness over both the polysilicon floating gate and the substrate. After adjusting the threshold of any enhancement devices, a second gate layer, of either polysilicon or a polysilicon/silicide sandwich, is deposited and selectively removed with the second oxide layer to define gates 95 and 97 for peripheral devices, as well as a second polysilicon gate 99 that, along with floating gate 71, forms a memory cell 30. Sources 105 and drains 107 are then formed using the polysilicon gates of the particular device as a self-aligning mask.

The process concludes by defining a double layer of conductive lines in the following manner. First, a boron/phosphorus-doped silica glass 121 covering is applied, contact holes 123 are etched, and the glass is heated to its flow temperature to round the corners of the contact holes. A first layer of conductive lines 131 is then defined. An insulative intermetal layer 133 is deposited, etched back and redeposited to form a substantially planar surface. Via holes 135 are wet/dry etched and the second layer of conductive lines 137 is then defined. A passivation layer 139 can be deposited over the second metal layer 137, or for single metal layer devices, over the first metal layer 131.

EEPROM 100 can program/erase faster if its coupling ratio can be made higher. Coupling ratio of memory cell 30 (and also of EEPROM 100) is the ratio of a first capacitance (not shown) formed between control gate 99 and floating gate 71 of cell 30 over a second capacitance (not shown) formed between floating gate 71 and p-substrate 40 of cell 30. The first and second capacitances are in series; therefore, when the coupling ratio of memory cell 30 increases, with other factors being the same, the voltage drop between floating gate 71

and p-subtrate 40 of cell 30 also increases. As a result, it is easier for electrons to tunnel through thin tunnel oxide layer 59 into floating gate 71. In other words, programming cell 30 becomes faster.

There are at least two methods to increase the coupling ratio of memory cell 30. A first method is to increase the first capacitance formed between control gate 99 and floating gate 71 of cell 30. One way to do this is to increase the overlapping area between control gate 99 and floating gate 71 of cell 30.

A second method is to decrease the second capacitance formed between floating gate 71 and p-substrate 40 of cell 30. This can be done by reducing the overlapping area between floating gate 71 and p-substrate 40 of cell 30. It should be noted that although increasing the thickness of a dedicated tunnel oxide region 59 between floating gate 71 and p-substrate 40 of cell 30 would decrease the second capacitance and hence increase the coupling ratio, that would also make it much harder for electrons to tunnel through the tunnel oxide region 59. Therefore, as a compromise, the dedicated tunnel oxide layer 59 should be thinner at only a small portion of the tunnel oxide region 130 to serve as a pathway for electrons to tunnel from p-substrate 40 into floating gate 71 and should be thicker at the rest of tunnel oxide region 59.

However, there is still room for improvement using the second method mentioned above. It is the object of the present invention to improve upon the prior art method of decreasing the second capacitance formed between the floating gate and the p-well or p-substrate, by providing a method of forming a memory cell in which the tunnel oxide region is thinner at a small portion in order to create a pathway for electrons to tunnel into the floating gate, while having the tunnel oxide region remain thicker at other places.

SUMMARY OF THE INVENTION

The non-volatile memory cell of the present invention has a small sidewall spacer electrically coupled and being located next to a main floating gate region. Both the small sidewall spacer and the main floating gate region are formed on a substrate and both form the floating gate of the non-volatile memory cell. Both are isolated electrically from the substrate by an oxide layer which is thinner between the small sidewall spacer and the substrate; and is thicker between the main floating gate region and the substrate. The small sidewall spacer can be made narrow; therefore, the thin portion of the oxide layer can also be made small to create a small pathway for electrons to tunnel into the floating gate.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a sectional view of a typical EEPROM (Electrically Erasable Programmable Read Only Memory) of prior art.

Figs. 2A-2H illustrate the steps in manufacturing the self-aligned non-volatile memory cell of the present invention.

Fig. 3 is another preferred embodiment of the non-volatile memory cell of the present invention.

BEST MODE FOR CARRYING OUT THE INVENTION

The final structure and operation of a non-volatile memory cell 200 of the present invention, as shown in Fig. 2H, can be well understood by going through the steps for fabricating the same. With reference to Fig. 2A, the fabrication process of the non-volatile memory cell of the present invention starts with, for illustrative purposes, a p-type semiconductor substrate 204. A silicon oxide (SiO2) layer 208, about 300Angstrom ($1Å=10^{-10}m$) thick, is formed on substrate 204. In the next step, a first polysilicon (poly-1) layer 212 is deposited upon silicon oxide layer 208. Then, excessive

portion of poly-1 layer 212 is etched away leaving only poly-1 region 212 as seen in Fig. 2A, which later serves as part of a floating gate 212, 239, 251 of memory cell 200 in Fig. 2H. In the next step, n+ regions 216 and 220 are implanted by ion bombardments using poly-1 region 212 as a mask. In other words, n+ regions 216 and 220 are self-aligned with the two opposing sides of poly-1 region 212.

With reference to Fig. 2B, photoresist masks 224 are used in wet-etching away a portion of silicon oxide layer 208 to expose surface 228 of n+ region 216. Then, masks 224 are removed.

With reference to Fig. 2C, a thin silicon oxide layer 232, about 70Å thick, is formed upon the structure, and covers completely the structure, including surface 228.

With reference to Fig. 2D, a second polysilicon (poly-2) layer 236 is deposited to blanket thin silicon oxide layer 232. Poly-2 layer 236 is then dry-etched away leaving only poly-2 sidewall spacers 239 and 242 to the left and right of poly-1 region 212, respectively, as seen in Fig. 2E. The etching operation can be carried out using an anisotropic etching method. Poly-2 sidewall spacer 242 to the right of poly-1 region 212 is not critical to the present invention, and therefore is not discussed in great detail in the following discussion. At this time, poly-2 sidewall spacer 239 and poly-1 region 212 are electrically separated by thin silicon oxide layer 232.

With reference to Fig. 2F, photoresist masks 245 are used in wet-etching away a portion of thin silicon oxide layer 232 to expose a surface 248 on top of poly-1 region 212. Then, masks 245 are removed.

With reference to Fig. 2G, a third polysilicon (poly-3) layer 251 is deposited to blanket the entire structure. This thin poly-3 layer 251 has electrical contact with poly-1 region 212 via surface 248. Poly-3 layer 251 also has direct contact with poly-2 sidewall

spacer 239. The excessive portion of poly-3 layer 251 is then dry-etched away leaving only the necessary portion 251 as seen in Fig. 2G.

With reference to Fig. 2H, an insulating ONO (Oxide/Nitride/Oxide) layer 254 is deposited to blanket the structure. Then, masks are used to remove the excessive portion of insulating ONO layer 254 to the left. In the next step, a fourth polysilicon (poly-4) layer 257 is deposited to blanket the entire structure. Then, masks are used to remove excessive portion on both sides of the poly-4 layer 257 as seen in Fig. 2H. The structure of insulating ONO layer 254 and poly-4 layer 257 and the method of fabricating the same are well known in the art, and therefore are not discussed in detail here.

The final structure is a non-volatile memory cell 200. Poly-1 region 212, poly-2 sidewall spacer 239, and poly-3 layer 251 form a floating gate 212, 239, 251 of memory cell 200. Poly-4 layer 257 forms a control gate 257 of memory cell 200. Floating gate 212, 239, 251 and control gate 257, separated from each other by insulating ONO layer 254 form a first parallel-plate capacitor (not shown). Floating gate 212, 239, 251 and substrate 204 forms a second parallel-plate capacitor (not shown). The first and second parallel-plate capacitors are in series. The insulating layer between the two parallel plates of the second parallel-plate capacitor has two insulating portions. A first insulating portion 260 is thin and is part of thin silicon oxide layer 232. First insulating portion 260 extends from the leftmost edge of poly-3 layer 251 to the rightmost edge of poly-2 sidewall spacer 239. A second insulating portion 263 is thicker and is part of silicon oxide layer 208 that lies under poly-1 region 212.

At a first look, both insulating portions 260 and 263 should be thick to keep low the capacitance of the second parallel-plate capacitor so as to keep high the coupling ratio of cell 200. However, such a high

coupling ratio would not make programming the cell easier because although most of voltage difference between control gate 257 and drain 216 would appear between floating gate 212, 239, 251 and drain 216 due to the high coupling ratio, it would still be difficult for electrons to tunnel through the thick insulating portions 260 and 263. Memory cell 200 of the present invention solves this problem by making insulating portion 260 thin and small. As a result, insulating portion 260 becomes a pathway (or tunneling oxide region) for electrons to tunnel from drain 216 into poly-2 sidewall spacer 239 which is part of floating gate 212, 239, 251 to program memory cell 200. Making insulating portion 260 thin increases the capacitance of the second parallel-plate capacitor. However, because insulating portion 260 is small in area compared with insulating portion 263, the increase of capacitance of the second parallel-plate capacitor is much less than if both insulating portions 260 & 263 are made thin to let electrons to tunnel into floating gate 212, 239, 251. As a result, this makes it easier for electrons to tunnel from drain 216 through thin insulating portion 260 into poly-2 sidewall spacer 239 which is part of floating gate 212, 239, 251 so as to program memory cell 200.

Programming memory cell 200 can be done by applying a high voltage (e.g., 12V-15V) to control gate 257, a ground potential to drain 216 and source 220. Electrons will tunnel through thin insulating portion 260 into poly-2 sidewall spacer 239 which is part of floating gate 212, 239, 251 under Fowler-Nordheim tunneling effect. The electrons trapped in floating gate 212, 239, 251 increases the threshold voltage of memory cell 200 such that there is no conducting channel between drain 216 and source 220 in read mode. In other words, a programmed cell 200 represents a logic 0.

An unprogrammed cell 200 having no trapped electrons in its floating gate 212, 239, 251 has a normal threshold voltage. In read mode, for an unprogrammed

cell 200, a conducting channel forms under insulating portion 263 between drain 216 and source 220. In other words, an unprogrammed cell 200 represents a logic 1. During read mode, the voltage applying to control gate 257 of memory cell 200 with respect to source 220 must be higher than the normal threshold voltage of an unprogrammed cell but must be lower than the increased threshold voltage of a programmed cell. As a result, during read mode, selected programmed cells 200 do not conduct and selected unprogrammed cells 200 conducts.

Erasing a programmed memory cell 200 can be done by applying a high voltage (e.g., 12V) to drain 216, a ground potential to both control gate 257 and source 220. Trapped electrons in floating gate 212, 239, 251 tunnels through thin insulating portion 260 to drain 216. Thereby, the cell becomes unprogrammed.

With reference to Fig. 3, another embodiment is shown in which memory cell 300 is the same as memory cell 200 of Fig. 2H, except that thin silicon oxide layer 232 is removed completely from the top of poly-1 region 212 by using a chemical-mechanical polishing (CMP) process to expose surface 248 of poly-1 region 212. After that, poly-3 layer 251, insulating ONO layer 254, and poly-4 layer 257 are in turn formed on the structure as in the case of memory cell 200 of Fig. 2H.

The non-volatile memory cell of the present invention also comprises a select transistor which is well known in the art and therefore is not discussed in here.